



下载APP



13 | 如何利用HRTF实现听音辨位？

2021-12-20 冯建元

《搞定音频技术》

课程介绍 >



讲述：冯建元

时长 11:02 大小 10.11M



你好，我是建元。

上节课，我们讲了空间音频的基本概念，以及空间音频是如何采集和播放的。相信你已经基本掌握了空间音频的基本原理。其实在游戏、社交、影视等场景中，空间音频被广泛地应用于构建虚拟的空间环境。

在空间音频的应用里最常见的一种就是“听音辨位”。比如在很多射击游戏中，我们能够通过耳机中目标的脚步、枪声等信息来判断目标的方向。



那么这节课我们就来看看，我们是如何利用 HRTF（Head Related Transfer Functions）头相关传递函数来实现“听音辨位”的。

HRTF 简介

上节课我们讲的“双耳效应”实际上就是空间中音源的声波从不同的方向传播到左右耳的路径不同，所以音量、音色、延迟在左右耳会产生不同的变化。

其实这些声波变化的过程就是我们说的声波的空间传递函数，是不是很耳熟？我们在讲回声消除的时候就是通过计算回声的空间传递函数来做回声信号估计的。

那么如果我们预先把空间中不同位置声源的空间传递函数都测量并记录下来，然后利用这个空间传递函数，我们只需要有一个普通的单声道音频以及这个音源和听音者所在虚拟空间中的位置信息，就可以用预先采集好的空间传递函数来渲染出左右耳的声音，从而实现“听音辨位”的功能了。

而这个和我们头部形状等信息相关的空间传递函数，也就是我们说的**头部相关传递函数（HRTF）**。

让我们再想一下：音源的声波是如何传递到我们的双耳的？一部分声音没有受到房间的墙壁、地板或者障碍物的干扰，而是直接通过空气传播到我们的双耳，我们把这些直接到达我们耳朵的声音叫做直达声。还有一些声波，经过空间障碍物或者界面的多次反射最后传播到你的耳朵里，从而形成了空间中的折射声或者说混响。很显然，直达声和混响相加就是我们听到的所有的声音了。

如果我们需要渲染一个真实的空间音频就需要渲染所有的直达声和混响。你有没有发现什么问题？

没错，直达声和虚拟环境是没有关系的，而混响则和听音者以及音源所处环境的空间布局有关。比如你在一个小房间和荒野大漠中的混响显然是不同的，而且混响还会和音源、听音者以及环境的相对位置有关，比如你在房间的墙角和在房间中央的混响信号，经过的路径显然是不同的。因此**直达声和混响我们通常需要分开来处理**。

那接下来就让我们先来看看 HRTF 中和直达声渲染相关的 HRIR（Head Related Impulse Response），也就是头部相关冲击响应是如何做直达声渲染的。

HRIR 与直达声渲染

HRIR 其实就是预先采集的直达声到达双耳的空间传递函数。为了得到一个线性系统的传递函数，我们通常会用一个脉冲信号作为系统的输入，然后记录这个系统的输出信号。当我们做直达声渲染时，只需要将原始音频卷积对应方向位置的双耳冲击响应就可以得到一个直达声的空间立体声了。

那么我们是如何采集 HRIR 的呢？

为了排除混响只录制 HRIR，我们需要在全消实验室录制 HRIR。图 1 就是一个在全消实验室录制 HRIR 的示意图。在图 1 中我们看到，全消实验室的墙壁上有很多吸音尖劈，这些材料可以有效地吸收声波，所以在全消实验室里墙壁是不会产生混响的。

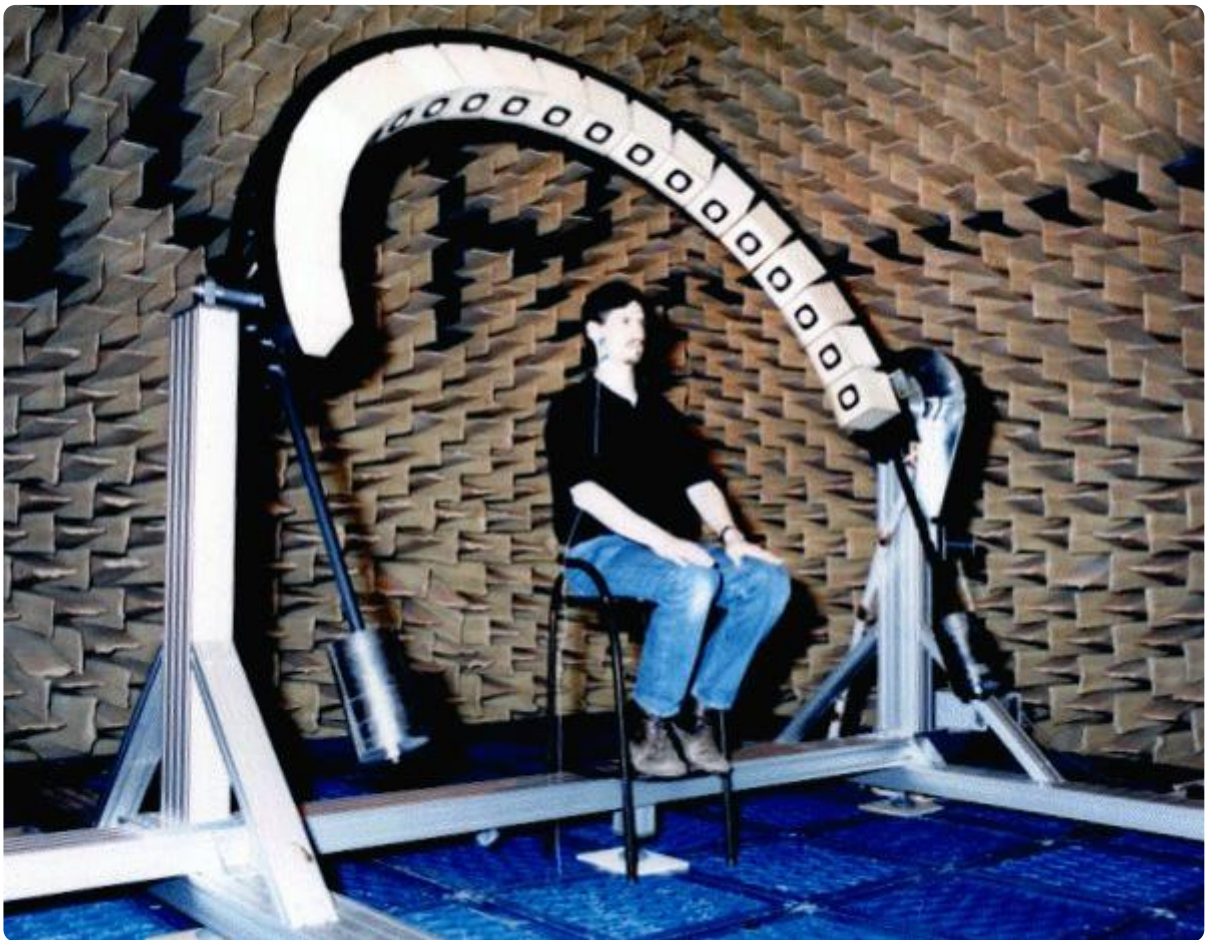


图1 全消实验室真人录制HRIR

在采集 HRIR 时，真人采集就是让人带上入耳式麦克风，然后让扬声器（白色的一圈）在一个球面的不同位置播放脉冲信号。这样入耳式麦克风采集到的就是空间中各个角度的 HRIR 了。这里的扬声器是一个半弧形阵列，测完一个半弧形后可以改变俯仰角再测一组，直至把所有角度都测完。

当我们把所有方向的数据都采集完毕后，就可以得到一个球形的数据集。图 2 就是一个球形 HRIR 的坐标示意图。我们可以看到图 2 中央是人头所在的位置，周围的红点就是 HRIR 采集时音源的方位。

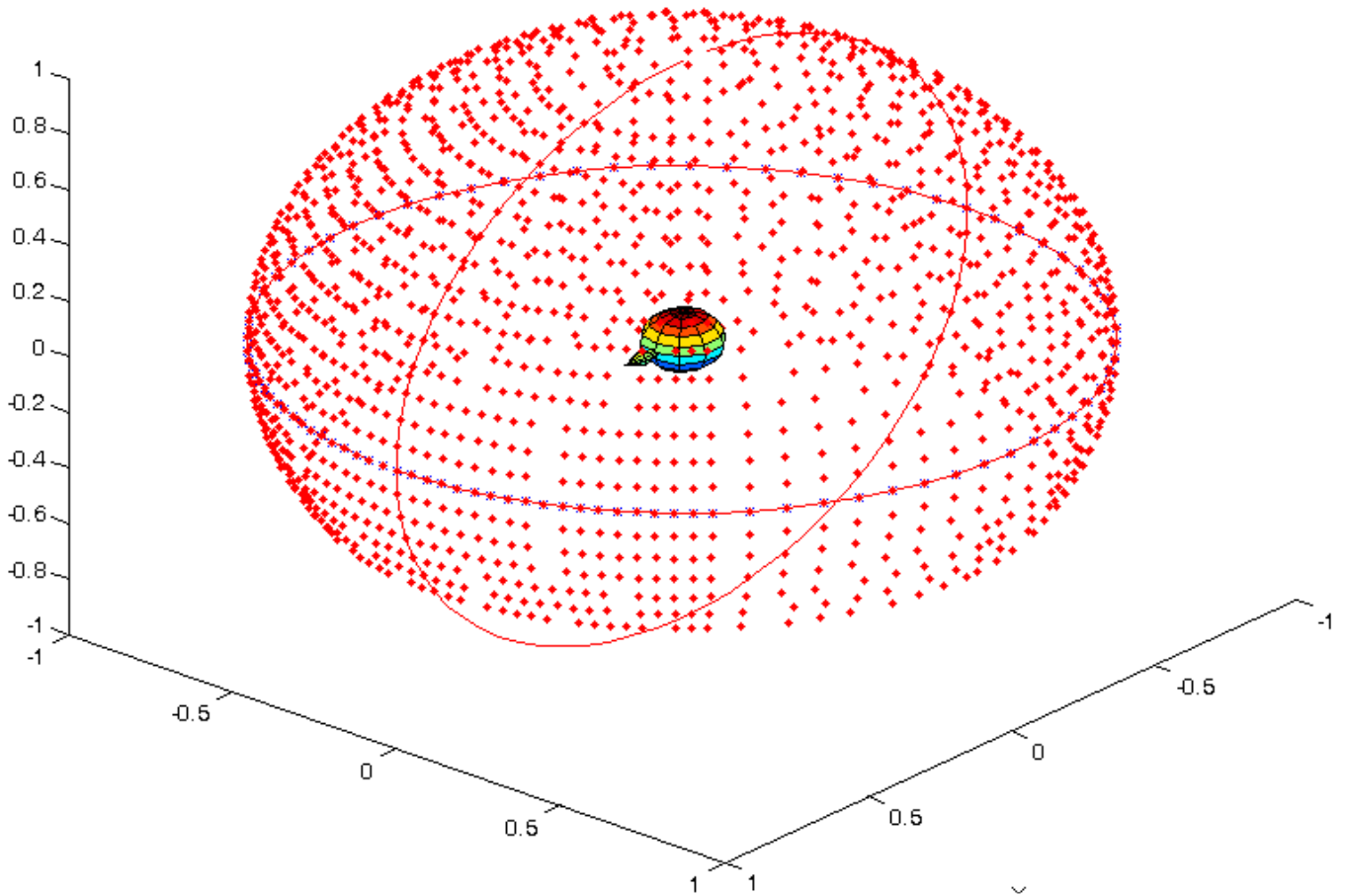


图2 HRIR的采集点的空间分布

当然让一个真人一动不动地测完全套可以得到比较准确的数据，但是这一整套流程下来时间还是很长的。所以为了方便采集，也可以采用同样的方法，使用人工头来采集 HRIR，比如图 3 就是一个利用人工头来采集 HRIR 的例子。人工头上的耳廓可以更换，这样就可以实现可定制耳廓的 HRIR 的采集。

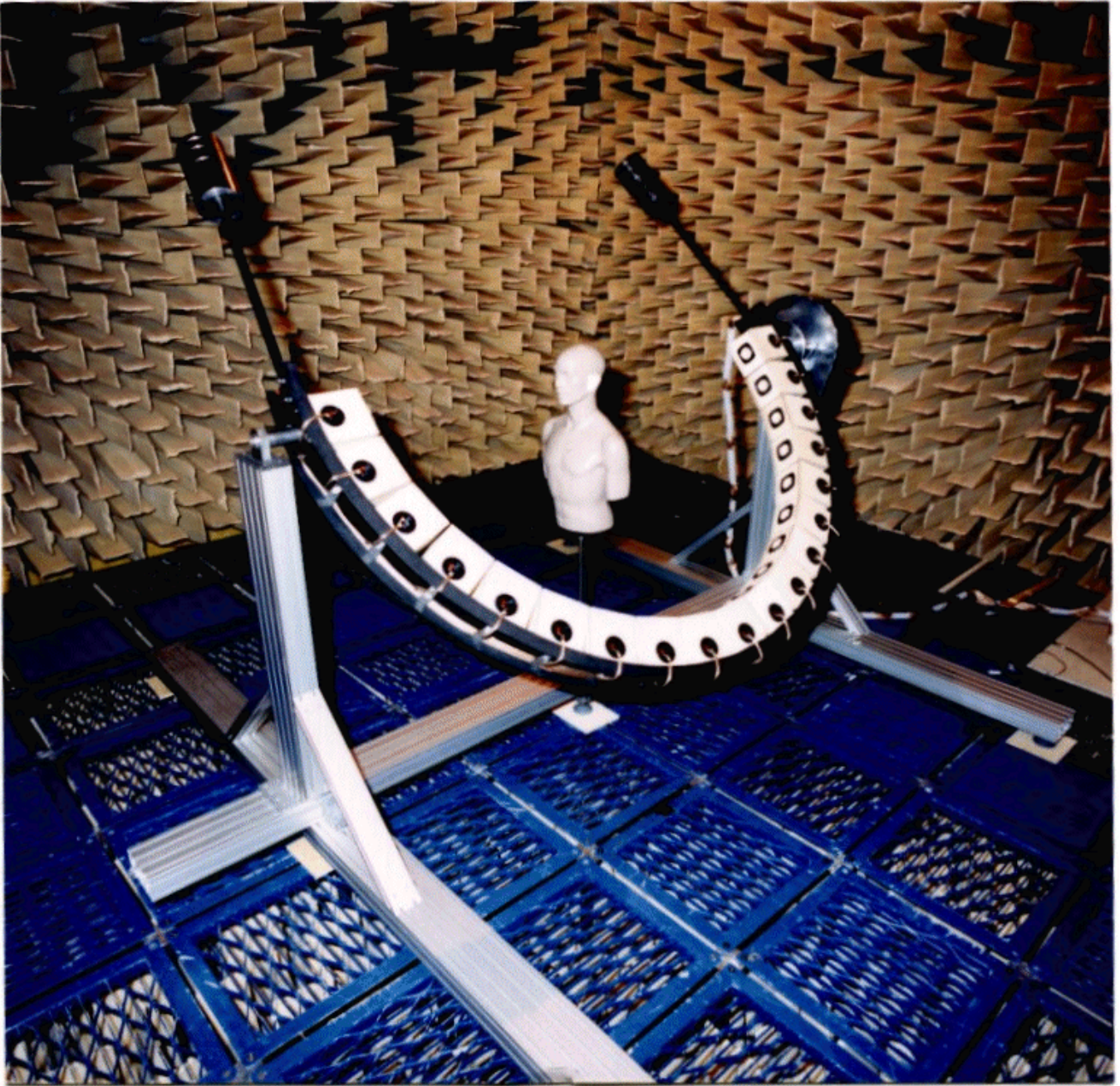


图3 全消实验室人工头录制HRIR

通过采集到的 HRIR，我们就可以把单通道的音频分别和某个角度的双耳 HRIR 做卷积，从而渲染出特定方向的声音。你可能会疑问，人的耳朵外形轮廓有的不太一样，这会不会影响空间音频的效果呢？

确实每个人的头部形状可能不尽相同，如果你用别人的 HRIR 来做空间音频的渲染可能没有你自己平时听得那么自然。从实际使用的结果上来看，**非定制化的 HRIR 渲染出的音频在角度的判断上差异不是很大，但是渲染后的音色可能会有一些细微的不同。**但目前定制化的 HRIR 成本还是比较昂贵的，也许在未来人们都进入“元宇宙”，这种定制化的增强听感的需求会催生出一些新的更便宜且准确的定制化方案。

好，那现在你知道 HRIR 是怎么采集的了，我再分享几个开源的 HRIR 的数据集。如果你有兴趣可以拿去试一试。一个是 [MIT 的人工头数据集](#)，一个是 [CIPIC 的真人数据集](#)。

知道了 HRIR 的采集和渲染方法之后，我们就可以对任意一个单声道的音频进行直达声方位的渲染了。那么剩下的混响我们要如何去渲染呢？

RIR 与反射声渲染

相较于直达声，反射声或者说混响和空间环境、音源位置、听音者位置都是绑定在一起的。所以如果采用和人工头相似的方法去采集房间冲击响应（RIR），那么理论上一个房间内音源位置和听音位置的组合可以说是无限多的。

并且就算可以用离散的方式采集一个房间内所有音源位置和听音位置的冲击响应，但是在虚拟环境里房间的形状、大小可以说是千变万化的，不可能每个都采用采样混响的方式来渲染。

在实际使用中，其实混响的作用不是提供准确的方位感而是提供空间大小，或者说听音者所处环境的感知。实际上根据镜像原理（如图 4 所示）一个真实声源的声波在经过界面反射后传到听音者的耳中可以等效成一个镜像声源的直达声。

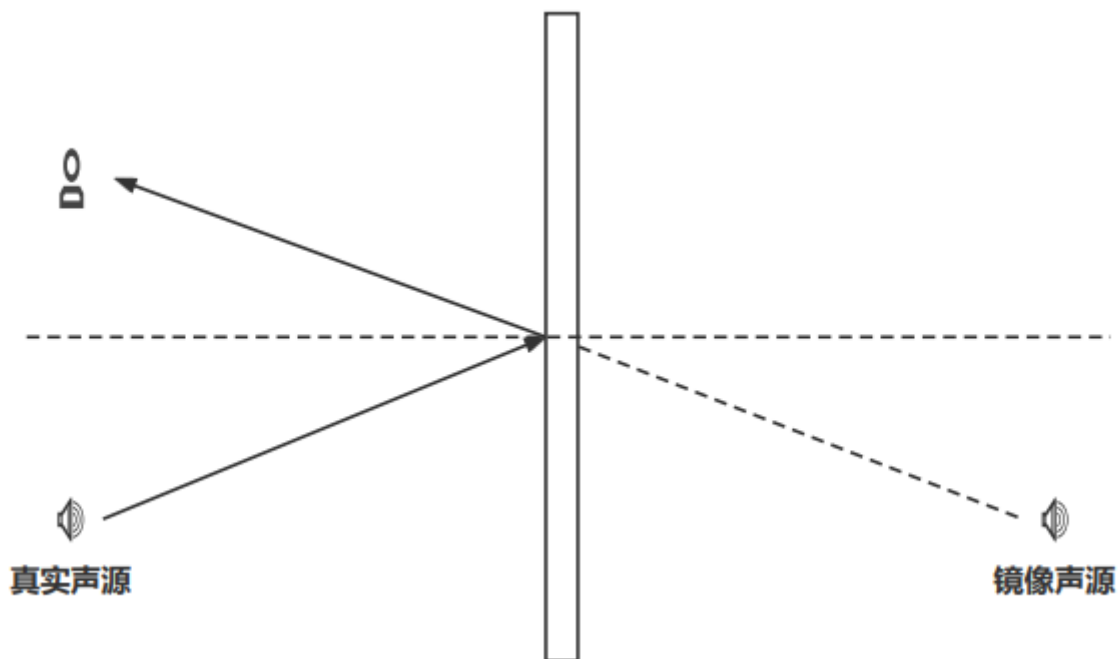


图4 折射声镜像原理

假设我们在一个长方体的房间里唱歌（如图 5 所示）。那么我们发出的声音经过 6 个界面的多次反射会形成多个镜像声源。所以其实混响会让对声源位置的判断变得更模糊，但是由于直达声一般来说会比经过反射的混响的音量大一些，所以在一般的房间内不至于让你的听音失去方向感。

这里值得注意的是，反射声大致可以分为前期反射和后期反射。我们一般把前 50ms 的反射叫做前期反射，超过 50ms 的叫做后期反射。一般来说前期反射对我们语音的可懂度是有提升效果的，而后期反射太多则会让声音变浑浊，从而降低语音的可懂度。

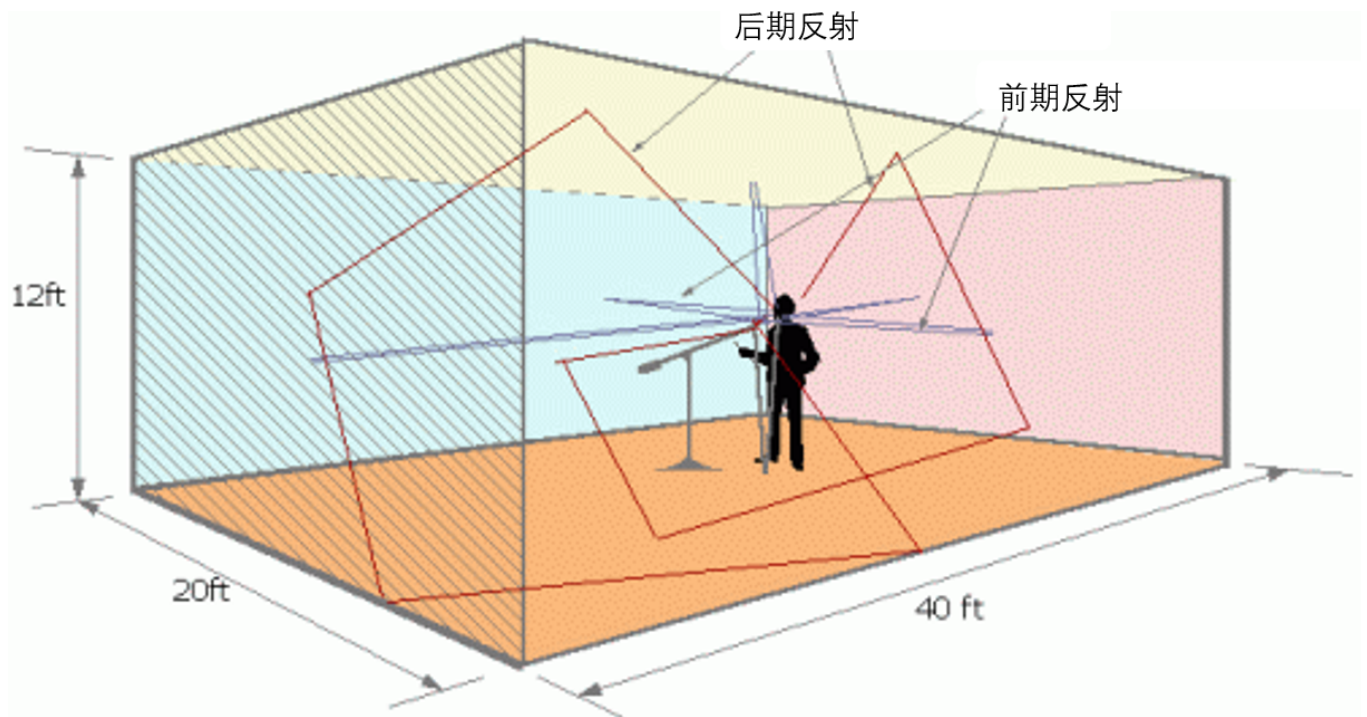


图5 长方体房间内的反射波示意图

这里讲的镜像原理其实也就是生成 RIR 的一种方法。这里我分享一个基于镜像法的 [房间冲击响应的生成器](#)，有兴趣的话你可以自行去了解一下。

其实镜像法的原理就是给定房间界面的大小、界面的反射吸音系数、空气衰减系数，再通过镜像法建模得到房间中任意位置音源和听音者位置组合的房间冲击响应。**用得到的冲击响应和原始音频做卷积就可以做混响的渲染了。**

在实际使用空间音频的时候，尤其是实时空间音频的渲染时，还有很多需要注意的地方，这里我罗列了几个常见的问题：

1. 如果我们想要用 RIR 给声音增加混响，那么我们原始的声音得是“干声”，也就是原始音频不能有混响。这一点是比较苛刻的。如果是在普通的房间录制的声音，比如客厅或者卧室，其实都已经有 200ms 到 1s 左右的混响了，再卷积一个 RIR 那么得到的混响可能比预期的更为浑浊，混响时间也更长一些。
2. 如果音源位置或者听音者位置是移动的，也就是说需要实时生成 RIR 来对混响进行建模，但镜像法生成混响，尤其在混响时间比较长的时候算力是比较大的。我们有的时候得做出取舍，比如一个房间就使用一个固定的预设 RIR 来避免算力无法实时计算。这也是很多游戏体验中混响在一个房间内都是一样的原因。
3. 真实的并不一定是好听的，在歌曲制作时我们经常会使用一些人工混响效果器来代替真实混响，或者采集一些比较好的固定混响的样本，比如用“维也纳金色大厅”的混响 RIR 来进行混响的渲染。

小结

空间音频渲染基本原理中的直达声原理和混响原理我们基本就介绍完了。利用这节课的知识你应该可以自己渲染出一个有方向和空间感的音频了。

现在我们来总结一下这节课的内容。

为了实现听音辨位，我们可以使用 HRTF 对空间音频进行渲染。

直达声可以采用预先采集的 HRIR 和原始音频通过卷积的方式来实现。基本上需要哪个方向的声音就去卷积哪个方向的 HRIR。HRIR 一般需要全消声实验室进行采集，我们可以使用一些开源库中的 HRIR 来实现。

反射声或者说混响的渲染则是采用卷积房间冲击响应 RIR 的形式来实现。由于 RIR 和房间的大小、材料、听音者和音源的位置都有关系，所以一般采用镜像法模拟的形式来实现。

在这里我还想补充一点，在实时交互的场景里，空间音频渲染时计算的实时性是很重要的。这里说的卷积和前面回声消除里讲的自适应滤波器一样，都是可以用频域卷积来加速的，比如采样率是 48kHz 的音频。如果需要和超过 64 点以上的卷积核做卷积，那么用频域卷积会快于时域卷积。

思考题

好了，今天的课程就到这里了，最后还有一个思考题留给你。我们如果想用耳机听音一个 5.0 声道的环绕立体声，我们应该如何用 HRTF 来实现？

你可以把你的答案和疑惑写下来，分享到留言区，与我一起讨论。我们下节课再见。

分享给需要的人，Ta订阅后你可得 **20** 元现金奖励

 生成海报并分享

 赞 0  提建议

© 版权归极客邦科技所有，未经许可不得传播售卖。页面已增加防盗追踪，如有侵权极客邦将依法追究其法律责任。

上一篇 12 | 空间音频入门：如何实现“声临其境”？

下一篇 14 | 音效三剑客：变调、均衡器、混响

精选留言

 写留言

由作者筛选后的优质留言将会公开显示，欢迎踊跃留言。